CRD report

News from the LBNL Computational Research Division — July 2004

They're Cool, but How Can We Use Them? Computation Sheds Light on New Nanostructures

Start with a quantum dot — a conglomeration of a few hundreds or thousands of atoms in which it is possible to control a single electron. Attach to the dot four rods of another material which are electronically coupled to the central quantum dot, then have the rods start branching. You're creating nanocrystals in various shapes that may have useful electronic or optical properties. But how do you find out what those properties are?

Ask the computational scientists on your team.

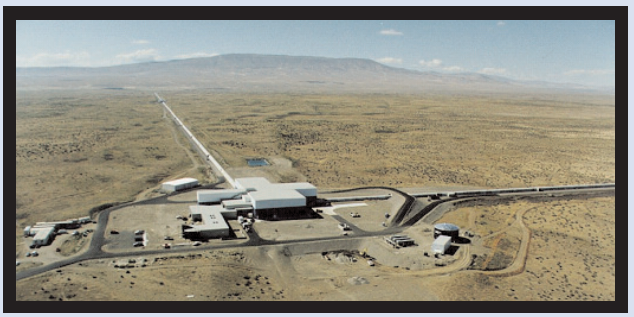
That's what Paul Alivisatos and his colleagues did. Alivisatos, director of Berkeley Lab's Materials Sciences Division and professor of chemistry at the University of California at Berkeley, and his research team have demonstrated a general approach for fabricating inorganically coupled colloidal quantum dots and rods. These remarkable new branched structures, whose dimensions are measured in billionths of a meter, are described in the July 8 issue of Nature [1].

Lin-Wang Wang and postdoctoral fellow Jingbo Li of CRD's Scientific Computing Group have worked closely with the Alivisatos group, using advanced computational techniques to identify the properties of new nanostructures coming out of the laboratory. Calculating the electronic properties of the structures depends on which compounds are involved, how they are joined, the number and arrangement of their atoms, and their proximity to other structures.

Analysis of the current batch of nanocrystals suggests that some of the new composite forms may have future applications in quantum computers, which operate by controlling the coherence (relatedness) of small collections of electron wavefunctions. And the unusual electronic band structures of other forms have possible implications for photovoltaic energy conversion in solar cells.

(continued on page 2)

Python/Globus Tools Speed Up Development of Data Grid for LIGO



The arms of the L-shaped LIGO observatories in Hanford, Washington and Livingston, Louisiana are 2.5 miles long. The long distance between the two facilities reduces the possibility of noise, such as seismic events, contaminating the data. Almost 1 TB of data is produced every day.

Programming tools developed by Keith Jackson and his colleagues in CRD's Secure Grid Technologies Group have been used to set up an efficient system to distribute new data that will put the predictions of Einstein's General Theory of Relativity to the test. To date, more than 50 TB of data from LIGO has been replicated to nine sites on two continents, quickly and robustly.

LIGO, the Laser Interferometer Gravitational-Wave Observatory, is a facility dedicated to detecting cosmic gravitational waves — ripples in the fabric of space and time — and interpreting these waves to provide a more complete picture of the universe. Funded by the National Science Foundation, LIGO consists of two widely separated installations — one in Hanford, Washington and the other in Livingston, Louisiana — operated in unison as a single observatory. Data from LIGO will be used to test the predictions of General Relativity — for example, whether gravitational waves propagate at the same speed as light, and whether the graviton particle has zero rest mass.

(continued on page 3)

China's NSF Honors Lin-Wang Wang with Outstanding Young Researcher Award

Lin-Wang Wang, a staff scientist in the Scientific Computing Group, recently won an Overseas Outstanding Young Researcher Award from the National Natural Science Foundation of China, China's equivalent of the U.S. National Science Foundation. These awards are given to recognize the recipients' academic achievements in their fields and to encourage collaborations with Chinese universities and research institutes.

The three-year award provides Lin-Wang with funding of about $50,000. He plans to use the funds to cover travel costs to China, to collaborate with Chinese scientists, and to fund student researchers.

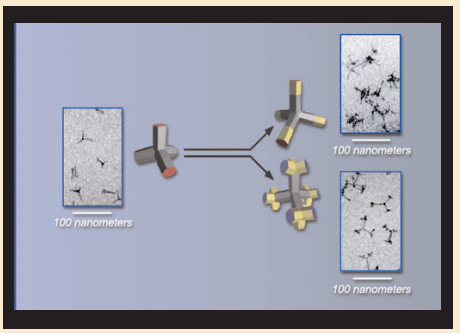
"We're going to use the methodologies I developed here at the Lab to study nanosystems, like the electronic structures of quantum dots," Wang said. "The students there can help us to improve these methods."

He will be working closely with the Semiconductor Institute in Beijing.

New Nanostructures (continued from p.1)

The Tricky In-Between Scale

Wang defines a nanostructure as "an assembly of building blocks on the scale where their properties become different from their bulk counterparts." A nanostructure may be composed of anywhere from a few hundred to a million atoms. While existing methods of calculation are well developed for very small collections of atoms, as in a molecule, or very large collections (virtually infinite) in bulk materials, the region in between is where calculations run into problems.



Examples of nanostructures made possible by the Alivisatos group's new method include tetrapods of cadmium selenide (left) extended with segments of cadmium telluride (upper right), which can be made to branch in turn (lower right).

"One would like to use *ab initio* methods, which start by simply inputting a few atomic numbers and deriving the properties and behavior of the system directly," Wang says. "The Schrödinger equation theoretically makes that possible, but in practice calculating anything larger than a single helium atom necessarily involves simplifications and approximations."

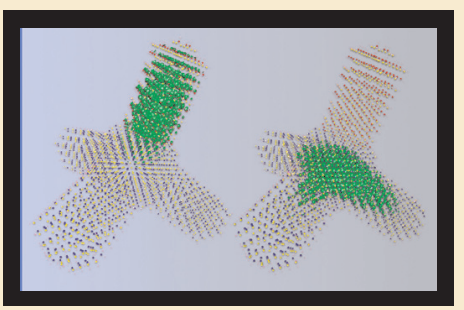
Physicists use a modified *ab initio* technique called the local density approximation (LDA) to calculate the electronic structures of small systems (density refers to the electron charge density in specific regions of the sample).

"Even with LDA, you could spend months doing a straightforward calculation of the electron wavefunctions and their charge densities for a nanosystem with thousands of atoms," Wang says. The memory needed for calculation increases as the square of the number of atoms in the system — while the needed processing power increases as the cube! A million-atom system is far out of reach of LDA.

So Wang and his colleagues developed a hybrid method, which draws on LDA to determine the charge density in one small region of a crystal, then by "charge patching" extends the charge density to the entire system, in this case a quantum dot or nanorod.

To accurately model a nanostructure, Wang begins by "passivating" its surface with fractionally charged pseudo-hydrogen atoms — mathematical entities that assign the surface atoms the same reactivity as that of a real nanoparticle in suspension. The positions of the interior atoms are calculated with the valence force field method, which models the strength, elasticity, and direction of bonds among atoms.

Then LDA is used to determine the charge "motifs" around a number of representative atoms, including the surface pseudo-hydrogens. Through charge patching, the calculation is extended to include the entire nanostructure. In a final step, a "folded spectrum" method that Wang developed 10 years ago is used to determine the material's electronic states near the band gap, including the highest-energy state of the valence band (which in an ideal semiconductor is filled with electrons) and the lowest energy state of the conduction band (which is empty of electrons).



NERSC calculations using local density approximation, the charge patching method, and the folded spectrum method yield atom-by-atom electronic maps of a tetrapod with one leg of cadmium selenide and three of cadmium telluride. On the left, green marks the conduction band's lowest energy state, which is physically separated in the structure from the valence band's highest energy state, shown in green on the right.

The various ways that compounds can be assembled into different structures yield very different optical and electronic properties. "The charge-patching method allows us to model thousand-atom structures with *ab initio* accuracy in about one hour, with the help of the Seaborg supercomputer at NERSC," says Wang. "This gives us unprecedented power to predict the electronic and optical properties of a given nanostructure."

Predicting and Tuning the Properties of Nanostructures

The marriage of nanostructure fabrication with the ability to precisely calculate electronic properties opens possibilities for collaboration in more ways than one. "In this case the researchers had already synthesized the structures, and we were able to tell them the electronic properties and how they change with alterations," Wang says. "In the future, by modeling a proposed system in advance, we could help decide what's possible and how to control their results."

The ability to independently tune the properties of each component and the nature of their interactions will enable researchers to create electronic devices tailored to a variety of uses. The structures can be chemically manufactured in large quantities, for potential applications ranging from ultrafast transistors to artificial photosynthesis.

Wang's research on nanoscale electronic structure calculations has attracted international attention (see story, p.1). While continuing to collaborate with the Alivisatos group, Wang is also making his expertise available to users of Berkeley Lab's new Molecular Foundry (foundry.lbl.gov). One of five DOE Nanoscale Science Research Centers, the Molecular Foundry is focused on the dissemination of techniques and methods for fabricating nanoscale patterned materials. As a computational science liaison to the Foundry's affiliated theory laboratory, Wang will help Foundry users apply LDA, empirical pseudopotential, and charge patching methods, and will continue working to develop new computational methods to meet the needs of the burgeoning fields of nanoscience and nanotechnology.

[1] Delia J. Milliron, Steven M. Hughes, Yi Cui, Liberato Manna, Jingbo Li, Lin-Wang Wang, and Paul Alivisatos, "Colloidal nanocrystal heterostructures with linear and branched topolo-

CRD report

July 2004

Find CRD Report and other CRD Presentations on the web at http://crd.lbl.gov/DOEresources

Summer Students Further Their Careers While Participating in Cutting-Edge Research

Each summer, Berkeley Lab's Computing Sciences organization hosts students from various universities in the United States and abroad. Additionally, through the DOE's Computational Science Graduate Fellowship (CSGF), which works to identify and support some of the best computational science graduate students in the nation, fellows participate in a three-month practicum at a DOE research laboratory. Berkeley Lab is hosting two CSGF fellows this summer.


Michael Wolf

One CSGF fellow at the Lab, Michael Wolf of the University of Illinois, Urbana-Champaign, is working on improving the parallel efficiency of the electromagnetic field solver Tau3P, which he co-developed while working at the Stanford Linear Accelerator Center (SLAC) for five years. While he was at SLAC he worked with Ali Pinar and Esmond Ng of the Scientific Computing Group (SCG), the same two researchers he is collaborating with this summer. His results of parallel mesh partitioning have been encouraging, and he also made some progress improving the matrix/vector multiplication algorithm used in Tau3P.

"When developing algorithms, there is a danger that you might forget about their application," said Michael, who's getting his Ph.D. in Computer Science. "I want to create algorithms that are useful to the scientific community, and labs such as LBNL afford me this opportunity."

With an undergraduate degree in biology and computer science from Harvey Mudd College, Michael finds he gravitates toward computational biology. During his Lab practicum he's also working on creating biological computational models with Physical Biosciences researcher Teresa Head-Gordon.

"Computational biology has these huge, seemingly unsolvable problems," he said. "I love the challenge of finding the balance between computational feasibility and model accuracy."


Ben Lewis

Computing Science's other CSGF fellow, Ben Lewis of the Massachusetts Institute of Technology (MIT), is also working in biology, specifically genomics. Ben is working with Mike Eisen in Life Sciences, comparing DNA sequences from multiple fruit fly genomes, and studying the evolution of DNA regions that may be involved in the transcriptional regulation of genes essential for normal fly development, metabolism and more. Regions of DNA that have been preserved in evolution among divergent species of fruit flies are more likely to be involved in processes that are essential to the organisms' fitness.

"High performance computing makes it possible to do thorough, detailed analyses of the vast amounts of genomic sequence data produced by several recent fruitfly genome projects," he said. "We split the genomes into little pieces and compare them in parallel; clusters allow us to do this quickly and efficiently."

Working in Eisen's lab, Ben is developing algorithms that make use of genome sequence data from six drosophila species to identify sequence motifs that may be involved in the transcriptional regulation of a family of genes called microRNA genes. The unique properties of microRNA genes, such as their small size and presumed importance in development, make these genes an ideal subject for computational comparative genomics analyses of gene expression.

(continued on page 4)

Python/Globus (continued from p.1)

Because gravitational waves have never been directly detected (although their influence on distant objects has been measured), LIGO is conducting blind searches of large sections of the sky and producing an enormous quantity of data — almost 1 TB a day — which requires large-scale computational resources for analysis.

The LIGO Scientific Collaboration (LSC) scientists at 41 institutions worldwide need fast, reliable, and secure access to the data. To optimize access, the datasets are replicated to computer and data storage hardware at nine sites: the two observatory sites plus Caltech, MIT, Penn State, the University of Wisconsin at Milwaukee (UWM), the Max Planck Institute for Gravitation Physics/Albert Einstein Institute in Potsdam, Germany, and Cardiff University and the University of Birmingham in the UK. The LSC DataGrid uses the DOEGrids Certificate Authority operated by ESnet to issue identity certificates and service certificates.

The data distribution tool used by the LSC DataGrid is the Lightweight Data Replicator (LDR), which was developed at UWM as part of the Grid Physics Network (GriPhyN) project. LDR is built on a foundation that includes the Globus Toolkit®, Python, and pyGlobus, an interface that enables Python access to the entire Globus Toolkit. LSC DataGrid engineer Scott Koranda describes Python as the "glue to hold it all together and make it robust."

pyGlobus is one of two Python tools developed by Jackson's group for the Globus Toolkit, the basic software used to create computational and data grids. The pyGlobus interface or "wrapper" allows the use of the entire Globus Toolkit from Python, a high-level, interpreted programming language that is widely used in the scientific and Web communities. pyGlobus is included in the current Globus Toolkit 3.2 release.

"What's great about using pyGlobus and Python is the speed and ease of development for setting up a new production grid application," Jackson said. "The scientists spend less time programming and move on to their real work — analyzing data — faster."

Another Python tool just released in the Globus Toolkit 3.9.2 Development Release (alpha test version for next year's GT 4.0) is the Python WS Core, a Python implementation of the Web Services Resource Framework (WS, RF) specifications. When GT 4.0 is released, the grid community will be moving from homegrown protocols and specifications to industry standard Web Service protocols for client and server support and secure messaging. Moving to the new standards will simplify the creation of Web services that can interface efficiently with many resources.

Jackson's development team for pyGlobus and the Python WS Core includes Joshua Boverhof, Noah Edelson, Monte Goode, David Konerding, David Robertson, and Matt Rodriguez. For more information about the Secure Grid Technologies Group, see <http://www-itg.lbl.gov/SGT/>, or contact Keith Jackson at <krjackson@lbl.gov>.

CRD report

July 2004

Find CRD Report and other CRD Presentations on the web at http://crd.lbl.gov/DOEresources

Summer Students (continued from p.3)

"Biology will be totally different in a few years," Ben said. "We're going through a transformation in how research is done. Computation-based discoveries are beginning to lead the way in research areas that have long been dominated by experimentalists."

In addition to hosting the CSGF fellows, Computing Sciences hosts its own Summer Student Program, which gives students an opportunity to gain relevant research experience while pursuing their degree. This year 16 students are partnering with one or more staff members on well-defined research projects.

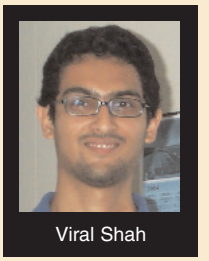
"We rely on a huge number of students," Deb Agarwal, head of the Distributed Systems Department, told the summer students at one gathering. "Don't underestimate your impact at the Lab; this is really important research."

Many of their projects, which they develop at the Lab during the 12-week summer program, become the basis for their theses. Several of the students presented their research findings in an open seminar on Tuesday, August 3. Here is a summary of their presentations:

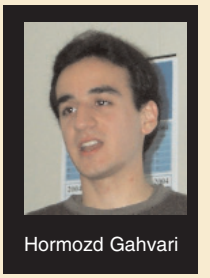

Ryan McKenzie

Ryan McKenzie from the University of Kentucky presented a talk titled "Building High-Level Tool Interfaces with Python." He addressed the advantages of using high-level software interfaces as teaching tools, specifically in the context of the DOE ACTS Collection. He also discussed the challenges in designing and implementing such an interface. Ryan has been working with SCG's Tony Drummond and Osni Marques.

Viral Shah from UC Santa Barbara presented "Parallel Programming without MPI." He is working on further developing Matlab*P, what he calls a "simpler, more elegant way to write parallel programming." He hopes that his work on Matlab*P will become part of his graduate thesis. Viral has been working with SCG's Parry Husbands, who wrote the original version of Matlab*P for his thesis.

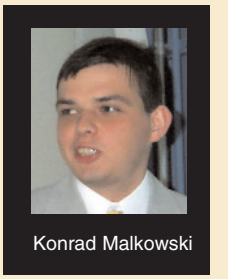

Viral Shah

Hormozd Gahvari from UC Berkeley presented "Benchmarking Sparse Matrix-Vector Multiplication." A sparse matrix is a matrix that contains mostly zeros with just a few non-zeros. The TOP500 supercomputers are benchmarked by Linpack, which is not a sparse matrix, but a dense matrix multiplication. Gahvari is interested in studying a new approach to benchmarking—creating a version of APEX Map that simultaneously runs on multiple streams. He works with Erich Strohmaier and Hongzhang Shan of the Future Technologies Group.

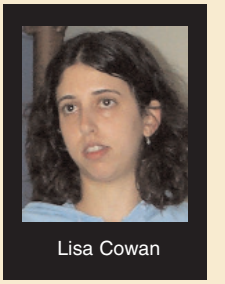

Hormozd Gahvari


Hui Xiong

Hui Xiong from the University of Minnesota presented "Hyperclique Pattern Discovery and Its Application to Protein Functional Module Extraction." Hui is researching how the application of a hyperclique pattern — a type of association pattern containing objects that are highly affiliated with each other — can identify functional modules in protein complexes. Proteins in the same functional module tend to be involved in common elementary biological functions. He has been working with SCG's Chris Ding and Stephen Holbrook in Physical Biosciences.


Konrad Malkowski

Konrad Malkowski from Pennsylvania State University presented "Data Mapping Techniques for Sparse Matrix Factorization." He is studying how to distribute data during sparse matrix factorization in order to improve system performance. He is also working on improving the accuracy of model's predictions. Specifically, he's studying the applications of multiple pass methodology in optimizing data mapping. The goal of his work is to make clusters more efficient during solution of sparse matrix systems using direct methods. Konrad has been working with SCG's Esmond Ng and Parry Husbands.


Lisa Cowan

Lisa Cowan from Mills College presented "Performance of Overlay Construction Algorithms in Representative Applications." An overlay network is a virtual network built on top of a physical network, such as the Internet. She is studying how to implement an overlay construction algorithm, implement it in a typical application and evaluate its performance. Lisa works with Karlo Berket in the Collaboration Technologies Group.

What is CRD Report?

This is CRD Report, a publication highlighting recent achievements by staff members in Berkeley Lab's Computational Research Division. Distributed every other month via email and posted on the Web at http://crd.lbl.gov/DOEresources, CRD Report may be freely distributed. CRD Report is edited by Jon Bashor, JBashor@lbl.gov or 510-486-5849.

DISCLAIMER
This document was prepared as an account of work sponsored by the United States Government. While this document is believed to contain correct information, neither the United States Government nor any agency thereof, nor The Regents of the University of California, nor any of their employees, makes any warranty, express or implied, or assumes any legal responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product, or process disclosed, or represents that its use would not infringe privately owned rights. Reference herein to any specific commercial product, process, or service by its trade name, trademark, manufacturer, or otherwise, does not necessarily constitute or imply its endorsement, recommendation, or favoring by the United States Government or any agency thereof, or The Regents of the University of California. The views and opinions of authors expressed herein do not necessarily state or reflect those of the United States Government or any agency thereof or The Regents of the University of California. Ernest Orlando Lawrence Berkeley National Laboratory is an equal opportunity employer.